

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-175092

(43) 公開日 平成11年(1999) 7月2日

(51) Int.Cl.⁶

識別記号

F I

G 1 0 L 3/00

5 5 1

G 1 0 L 3/00

5 5 1 G

G 0 9 F 27/00

G 0 9 F 27/00

N

H 0 4 N 5/278

H 0 4 N 5/278

審査請求 未請求 請求項の数19 OL (全 13 頁)

(21) 出願番号 特願平10-272969

(22) 出願日 平成10年(1998) 9月28日

(31) 優先権主張番号 08/938735

(32) 優先日 1997年9月26日

(33) 優先権主張国 米国 (U S)

(71) 出願人 595097771

アドビ システムズ, インコーポレイテッド

ADOBE SYSTEMS, INC.

アメリカ合衆国, カリフォルニア 95110,

サン ノゼ, パーク アベニュー 345

(72) 発明者 ビーター エフ. ウルマン

アメリカ合衆国, カリフォルニア

95112, サン ノゼ, サウス サード

ストリート 114, ナンバー 310

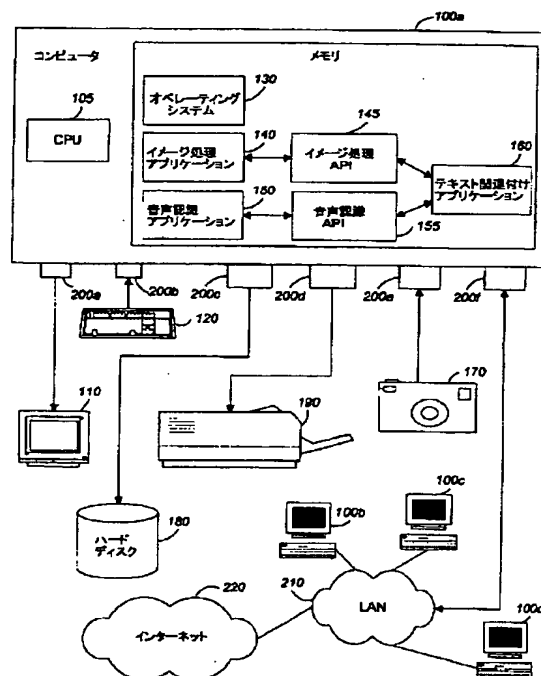
(74) 代理人 弁理士 小橋 一男 (外 1 名)

(54) 【発明の名称】 オーディオから派生したテキストのイメージとの関連付け

(57) 【要約】

【課題】 オーディオから変換させたテキストをイメージと共に表示させる方法、システム及び装置を提供する。

【解決手段】 イメージが感光性フィルムカメラ又はデジタルカメラを使用してキャプチャされ、又はコンピュータグラフィックスソフトウェアを使用して作成される。オーディオがイメージのキャプチャと同時に又は別の時間のいずれかにキャプチャされる。これらのキャプチャされたイメージ及びオーディオは格納され且つマルチメディアファイルフォーマットを使用して互に関連付けられる。オーディオは音声認識ソフトウェアを使用してテキストへ変換される。その変換されたテキストを該イメージ上又はその近くに位置決めすることによって該イメージと該変換したテキストから複合画像を形成する。その複合画像をコンピュータモニタ、プリンタ又はその他の出力装置へ出力する。



【特許請求の範囲】

【請求項1】 テキストと共にイメージを表示するためのコンピュータによって実行される方法において、コンピュータによって読取可能なメモリ内にオーディオデータとイメージデータとを与え、前記オーディオデータに対して計算音声・テキスト変換プロセスを適用して変換させたテキストを発生し、前記イメージデータを前記変換させたテキストと合成させることにより複合イメージを作成し、前記複合イメージを表示する、ことを特徴とする方法。

【請求項2】 請求項1において、前記表示を行なう場合に、前記複合イメージを印刷することを特徴とする方法。

【請求項3】 請求項2において、前記オーディオデータ及び前記イメージデータが好適にはフラッシュピクス(FlashPix)ファイルである単一のソースファイルのコンポーネントであることを特徴とする方法。

【請求項4】 請求項2において、前記オーディオデータ及び前記イメージデータが別々のソースファイルから発生することを特徴とする方法。

【請求項5】 請求項2において、前記コンピュータによって読取可能なメモリがコンピュータプログラム命令を実行するために動作可能なコンピュータのランダムアクセスメモリであり、前記計算音声・テキスト変換プロセスが前記コンピュータ上で実行されるコンピュータプログラム命令を有しており、前記コンピュータによって読取可能なメモリ内にオーディオデータとイメージデータとを与えるステップが、前記オーディオデータと前記イメージデータとを互いにリンクさせる情報を得、大量格納装置に格納されている前記イメージデータ及び前記オーディオデータを見つけ出し、且つ前記大量格納装置から前記ランダムアクセスメモリ内に前記イメージデータ及び前記オーディオデータを読み込むことを包含しており、前記複合画像を作成するステップが、前記イメージデータと前記変換したテキストとをイメージ処理アプリケーションプログラム内へインポートし、前記イメージ処理アプリケーションプログラムを実行して前記変換させたテキストをフォーマットし、前記テキストを前記イメージに関連して配置させ、且つ前記フォーマットし且つ配置させたテキストを前記イメージデータと合成させて前記複合画像を発生させることを包含している、ことを特徴とする方法。

【請求項6】 請求項5において、前記イメージデータ及び前記オーディオデータが前記大量格納装置に格納されている別々のファイル内に存在しており、且つ前記オーディオデータと前記イメージデータとを互いにリンクさせる情報が前記別々のファイルのうちの少なくとも1つに格納されているタグであることを特徴とする方法。

【請求項7】 請求項5において、前記変換させたテキストが前記イメージデータによって表わされるイメージの一部をカバーするように合成されることを特徴とする方法。

【請求項8】 請求項5において、前記変換されたテキストが、前記イメージデータによって表わされるイメージのいずれの部分もカバーすることがないように合成されることを特徴とする方法。

【請求項9】 請求項5において、前記イメージデータが単一のイメージを表わすことを特徴とする方法。

【請求項10】 請求項5において、前記イメージデータが複数の単一のイメージのシーケンスを表わしており且つ前記オーディオデータが複数のオーディオセグメントのシーケンスを表わしており、本方法が、更に、前記オーディオセグメントのシーケンスのうちの1つのオーディオセグメントを前記単一イメージのシーケンスのうちの1つの単一イメージとマッチングさせ、前記1つのオーディオセグメントを変換されたテキストセグメントへ変換し、

前記1つの単一イメージと前記変換させたテキストセグメントとを合成することにより単一の複合イメージを作成する、ことを特徴とする方法。

【請求項11】 請求項2において、前記イメージデータが複数の単一イメージからなるシーケンスを表わしており、本方法が、更に、前記単一イメージのシーケンスのうちの各単一イメージに対して、前記単一イメージと前記変換したテキストとを合成することによって複合イメージを作成し、且つ前記複合イメージの各々を印刷する、ことを特徴とする方法。

【請求項12】 請求項11において、前記カメラは、マイクロフォンを有しており且つ音声を記録し且つ記録した音声を前記カメラによってとったイメージと関連付けるべく動作可能なデジタルカメラであることを特徴とする方法。

【請求項13】 請求項1において、更に、前記コンピュータプログラム命令を実行すべく動作可能なコンピュータのランダムアクセスメモリ内にイメージデータとオーディオデータとを読み込み、前記コンピュータ上で実行されるコンピュータプログラム命令を包含する計算音声・テキスト変換プロセスを前記ランダムアクセスメモリ内の前記オーディオデータに対して適用して変換させたテキストを発生し、前記イメージデータと前記変換させたテキストとをイメージ処理アプリケーションプログラム内にインポートし、前記イメージ処理アプリケーションプログラムを実行して前記変換させたテキストをフォーマットし、前記テキストを前記イメージの底部中央近くに配置させ、且つ前記フォーマットし中心位置決めさせたテキストを前記イメージデータと合成して複合画像を発生し、

前記複合画像を印刷する、ことを特徴とする方法。

【請求項14】 コンピュータによって読取可能な媒体に格納されているコンピュータプログラム製品において、オーディオデータとイメージデータとをコンピュータによって読取可能なメモリ内に読み込ませ、前記オーディオデータをテキストへ変換させ、前記イメージデータと前記変換したテキストとを合成させて、複合画像を作成し、好適には前記複合画像を印刷することにより前記複合画像を表示させる、上記各命令を有することを特徴とするコンピュータプログラム製品。

【請求項15】 テキストと共にイメージを表示させるシステムにおいて、オーディオデータとイメージデータとをコンピュータによって読取可能なメモリ内に読み込む手段、前記オーディオデータをテキストへ変換させる手段、複合イメージを作成するために前記イメージデータと前記変換させたテキストとを合成する手段、好適には前記複合イメージを印刷することにより前記複合画像を表示する手段、を有することを特徴とするシステム。

【請求項16】 請求項15において、更に、コンピュータプログラム命令を実行すべく動作可能なコンピュータのランダムアクセスメモリ内にイメージデータとオーディオデータとを読み込む手段、前記コンピュータ上で実行されるコンピュータプログラム命令を有する計算音声・テキスト変換プロセスを変換させたテキストを発生させるために前記ランダムアクセスメモリ内の前記オーディオデータに対して適用させる手段、イメージ処理アプリケーションプログラム内に前記イメージデータと前記変換させたテキストとをインポートし、前記変換させたテキストをフォーマットし、前記テキストを前記イメージの底部中心近くに配置させ、且つ前記フォーマットし中心位置決めさせたテキストを前記イメージデータと合成させて複合画像を発生させるために前記イメージ処理アプリケーションプログラムを実行する手段、を有することを特徴とするシステム。

【請求項17】 請求項15において、前記コンピュータによって読取可能なメモリがコンピュータプログラム命令を実行すべく動作可能なコンピュータのランダムアクセスメモリであり、本システムが、更に、コンピュータプログラム命令を実行すべく動作可能なコンピュータのランダムアクセスメモリ内にイメージデータとオーディオデータとを読み込む手段、変換させたテキストを発生させるために前記ランダムアクセスメモリ内の前記オーディオデータに対して前記コンピュータ上で実行されるコンピュータプログラム命令を有する計算音声・テキスト変換プロセスを適用する手

段、

前記オーディオデータと前記イメージデータとを互いにリンクさせる情報を得、大量格納装置に格納されている前記イメージデータと前記オーディオデータとを見つけ出し、且つ前記大量格納装置から前記ランダムアクセスメモリ内に前記イメージデータと前記オーディオデータとを読み込む手段、

前記イメージデータと前記変換されたテキストとをイメージ処理アプリケーションプログラム内にインポートし、前記変換させたテキストをフォーマットし、前記テキストを前記イメージに関して配置させ、前記フォーマットし配置させたテキストを前記イメージデータと合成して前記複合画像を発生させるために前記イメージ処理アプリケーションプログラムを実行する手段、を有することを特徴とするシステム。

【請求項18】 請求項15において、前記イメージデータが複数の単一イメージからなるシーケンスを表わしており且つオーディオデータが複数のオーディオセグメントからなるシーケンスを表わしており、本システムが、更に、

前記複数のオーディオセグメントからなるシーケンスのうちの1つのオーディオセグメントを前記複数の単一イメージからなるシーケンスのうちの1つの単一イメージとマッチングさせる手段、

前記1つのオーディオセグメントを変換させたテキストセグメントへ変換する手段、

前記1つの単一イメージを前記変換させたテキストセグメントと合成することによって単一複合画像を作成する手段、を有することを特徴とするシステム。

【請求項19】 請求項15において、前記イメージデータが複数の単一イメージからなるシーケンスを表わしており、本システムが、更に、

前記単一イメージと前記変換させたテキストとを合成することによって前記複数の単一イメージからなるシーケンスの各単一イメージに対して複合イメージを作成する手段、

前記複合イメージの各々を印刷する手段、を有することを特徴とするシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、テキストをイメージと関連付けさせる技術に関するものであって、その場合に該テキストは該イメージと関連しているオーディオから派生させるものである。

【0002】

【従来の技術】カメラはイメージをキャプチャし且つそれらをフィルム又はデジタル媒体上に格納する。イメージをキャプチャした時と、それを印刷又はその他の態様で表示させる時との間において、写真家は例えばそれをキャプチャした時又はその主題となった位置等のそのイ

メージに関連する情報を忘れたり又はその情報に対するアクセスを失う場合がある。

【0003】幾つかのフィルムカメラ及びデジタルカメラは、例えばイメージをキャプチャした日付を表わすテキスト等のテキスト又はデジタル情報を写真と関連づけさせることを可能とする。このテキストは、典型的に、そのカメラによって作成され且つ所定のフォーマットで所定の位置においてそのイメージ上に重ね合わされる。

【0004】デジタルカメラはイメージをキャプチャし且つそれをコンピュータによって読取可能なディスク、フラッシュメモリ、又はその他の持続性のある格納（記憶）媒体上にデジタル形態で格納する。各イメージはスタンダードのフォーマットに従って別々のファイル内に格納することが可能である。次いで、該ファイルをコンピュータのメモリへ転送し、そこで、それをコンピュータソフトウェア又はハードウェアによって処理することが可能である。

【0005】オーディオ及びその他の情報はイメージファイル即ち画像ファイルと関連づけさせることが可能である。例えば、コダック社のフラッシュピックス（FlashPix）イメージ仕様は、デジタルカメラでキャプチャしたイメージ（画像）を格納するためのスタンダードのファイルフォーマットを特定している。このフラッシュピックスの仕様に対する拡張は、1つ又はそれ以上のオーディオストリームを1つのイメージファイルと関連づけさせ、従ってそれと共に格納させることを可能とする。この拡張に従って動作するソフトウェアは、イメージファイルを表示させながら、そのイメージファイルと関連するオーディオストリームのうちの1つ又はそれ以上をプレイバックすることを可能とする。

【0006】音声認識ソフトウェアは人間の音声を表わすオーディオ信号をテキストへ変換する。音声認識ソフトウェアは制限させたタスクに特化させることが可能であり且つ制限した数の単語を認識することを可能であり、又はより一般的なものとし且つ音声を音声学的に種類分けすることによってテキストを作成することが可能である。音声認識ソフトウェアはデジタル的に表わされたオーディオからコンピュータによって読取可能なテキストを作成することが可能である。そのようにして作成されたテキストは、コンピュータソフトウェアによって解釈し且つ処理を行ない且つコンピュータによって読取可能な媒体上に格納することが可能である。

【0007】写真を撮影した場合に関連する情報を記録するためにカメラと共にテープレコーダ又はメモ帳を持ち歩くことによって写真を撮影した時にオーディオ又はテキストを写真と関連付けさせることが可能である。幾つかのデジタルカメラでは撮影をしている間に画像と共にオーディオを直接的に記録することを可能とする。次いで、その画像を見る場合にオーディオをプレイバックすることが可能である。

【0008】

【発明が解決しようとする課題】本発明は、以上の点に鑑みなされたものであって、オーディオから変換させたテキストを画像と共に表示させることを可能とした技術を提供することを目的とする。

【0009】

【課題を解決するための手段】1側面においては、本発明は、変換させたテキストを発生させるためにコンピュータによって読取可能なメモリ内のオーディオデータに対して計算音声・テキスト変換プロセスを適用する技術を提供している。コンピュータによって読取可能なメモリ内に格納されているイメージデータを該変換させたテキストと合成することによって複合イメージが作成される。次いで、この複合イメージをプリンタ又はその他の適宜の出力装置において印刷する。該イメージデータは、カメラによって撮影したイメージによって表わされるものとすることが可能であり且つ該オーディオデータはそのイメージを撮影したのとはほぼ同時に記録した音声を表わすものとすることが可能である。該カメラは、マイクロフォンを有しており、且つ音声を記録し且つ記録した音声を該カメラによって撮影したイメージと関連付けすべく動作可能なデジタルカメラとすることが可能である。

【0010】該オーディオデータ及びイメージデータは、単一のソースファイルのコンポーネントとすることが可能であり、該ファイルはフラッシュピックスのフォーマットにおける1つのファイルとすることが可能である。該オーディオデータ及びイメージデータは、別々のソースファイルからのものとすることが可能である。該オーディオデータ及びイメージデータは、該オーディオデータと共に又は該イメージデータと共に又は該オーディオデータとイメージデータの両方と共に格納されているタグによってリンクさせることが可能である。該変換されたテキストは、該オーディオデータに対するエイリアスとして大量格納装置に格納することが可能である。該変換されたテキストは、該イメージデータによって表わされるイメージの一部をカバーするように、又は該イメージデータによって表わされるイメージのいずれの部分もカバーすることがないように合成させることが可能である。

【0011】該イメージデータは複数個の単一イメージからなるシーケンスを表わすことが可能であり、且つ該オーディオデータは複数個のオーディオセグメントからなるシーケンスを表わすことが可能である。複数個のオーディオセグメントからなるシーケンスのうちの1つのオーディオセグメントは、該1つのオーディオセグメントを変換させたテキストセグメントへ変換し且つ該1つの単一イメージと該変換させたテキストセグメントとを合成することによって単一の複合イメージを作成する前に、複数個の単一イメージからなるシーケンスのうちの

1つの単一イメージとマッチングさせることが可能である。一方、該複数個の単一イメージからなるシーケンスの各単一イメージに対して、該単一イメージを該変換させたテキストと合成することによって1個の複合イメージを作成することが可能である。本発明の1つの利点は、それがオーディオデータから派生されたテキストをイメージデータと関連付けさせることである。テキストは、典型的に、オーディオデータよりも一層迅速に且つ効率的に格納され、転送され且つ処理させることが可能なものであるから、オーディオデータをテキストへ変換し且つ該変換させたテキストをイメージと関連付けさせることは、イメージと共にオーディオを格納する場合と比較して有益的なものである場合がある。

【0012】本発明の別の利点としては、それがオーディオデータを変換させたテキストへ変換させ且つ該変換させたテキストを該イメージデータと共に印刷させることによって、イメージデータが印刷される場合、又は音声の再生が可能でないか又は実際的なものでない場合に、オーディオデータとイメージデータとの関連付けが持続することを可能とするということである。このように、本発明は、ユーザによる付加的な努力を必要とすることなしに、イメージデータに関連しているオーディオデータ内に含まれているような印刷した記述的情報を提供する。

【0013】本発明の更に別の利点は、変換させたテキストはスタンダードのコンピュータテキストフォーマットで格納することが可能であり、従ってスタンダードのコンピュータテキスト処理ソフトウェアによって処理及びフォーマットすることが可能である。

【0014】本発明の更に別の利点は、イメージデータ及びそれと関連するオーディオデータを同時に又は異なる時間にキャプチャすることが可能であるという点である。それらは、例えばマイクロフォン及びオーディオキャプチャ能力を有するデジタルカメラ等の単一の装置を使用して、又は異なる装置によってキャプチャすることが可能である。イメージデータ又はオーディオデータのいずれか一方、又は両方を、人間が発生させたもの又は人工的に作成したものとすることが可能である。従って、本発明は、イメージ、オーディオ、テキストを作成し、処理し、且つ格納する上での柔軟性を提供している。

【0015】

【発明の実施の形態】図1を参照すると、コンピュータ100aはI/Oポート200aにおいて接続しているモニタ110上において出力を表示し、I/Oポート200bにおいて接続しているキーボード120からの入力を得、且つI/Oポート200dにおいて接続しているプリンタ190上にハードコピーを出力する。コンピュータ100aは、更に、I/Oポート200cにおいてファイル及びその他のデータを格納し且つ検索するた

めのハードディスク180へ接続しており、他のコンピュータ100b-dと通信するためにLAN210へ接続しており、且つ情報を格納し、アーカイブし、検索するためにLAN210を介してインターネット200へ接続している。デジタルカメラ170はスチールイメージ即ち静止画像をキャプチャ即ち撮影することが可能である。デジタルカメラ170は、更に、イメージと関連するオーディオデータをキャプチャ即ち採取するためのマイクロフォンを有している。デジタルカメラ170は、これらのイメージデータ及びオーディオデータをデジタルフォーマットへ変換し、それらをたとえばフラッシュピククスフォーマット等のオーディオコンポーネント及びイメージコンポーネントを有するマルチメディアファイルフォーマットで格納する。そのファイルはI/Oポート200eを介してコンピュータ100aへ転送される。

【0016】アドビアクロバット(Acrobat)プログラム(カリフォルニア州サンノゼのアドビシステムズインコーレイテッドから入手可能)のようなイメージ処理アプリケーション140がコンピュータ100a上で稼動される。コンピュータ100a上においては、更に、音声を表わすオーディオデータを変換させたテキストへ変換し且つその変換させたテキストをコンピュータによって読取可能なファイル内に格納することが可能な例えばドラゴンシステムのドラゴンディクテイト(Dragon Dictate)のような音声認識アプリケーション150が稼動される。更に、コンピュータ150a上においては、変換されたテキストをイメージと関連付けさせるためのテキスト関連付けアプリケーション160が稼動される。テキスト関連付けアプリケーション160は、イメージ処理API145を介してイメージ処理アプリケーションと通信を行ない、それは、イメージ、オーディオ、テキストをイメージ処理アプリケーション140のファイル内にインポートする手順を包含している。テキスト関連付けアプリケーション160は、音声認識APIを介して音声認識アプリケーション150と通信を行ない、それはオーディオデータを変換されたテキストへ変換し且つ変換されたテキストをテキストファイル内に保存するための手順を包含している。

【0017】テキスト関連付けアプリケーション160はユーザに対して多数の選択可能なオプションを提供する。そのようなオプションのうちの1つは、例えばフラッシュピククスファイル等のソースファイルを印刷することである。ソースファイルは、最少で、イメージコンポーネントを有するものであるか、又、オーディオコンポーネントも包含するか又はそれと関連付けさせることが可能である。図2を参照すると、ユーザが、ソースファイルを印刷するためのオプションを選択すると(ステップ300)、テキスト関連付けアプリケーション160は、オーディオコンポーネントがソースファイル内に

包含されているイメージコンポーネントと関連付けされているか否かを判別する(ステップ310)。オーディオコンポーネントがイメージコンポーネントと関連付けされていない場合には、テキスト関連付けアプリケーション160は、イメージ処理API145のファイルを開く機能呼び出すことによってイメージ処理アプリケーション140におけるソースファイルを開く(ステップ320)。次いで、テキスト関連付けアプリケーション160は、イメージ処理API145の印刷機能呼び出すことによって該イメージコンポーネントを印刷する(ステップ370)。

【0018】オーディオコンポーネントがソースファイル内に含まれているか又はソースファイルと関連している場合には、ユーザは、そのオーディオコンポーネントを変換されたテキストへ変換するためのオプションが与えられる(ステップ330)。ユーザがそのオプションを辞退する場合には、イメージコンポーネントが開かれ且つ印刷される(ステップ320及び370)。ユーザがそのオプションを採用する場合には、テキスト関連付けアプリケーション160はそのオーディオコンポーネントを探し出す(ステップ340)。次いで、テキスト関連付けアプリケーション160は、音声認識アプリケーションAPI155の音声・テキスト変換機能呼び出すことによってそのオーディオコンポーネントを変換されたテキストへ変換する(ステップ350)。音声認識アプリケーション150は、変換されたテキストをファイル内に格納し且つそのファイルの名前をテキスト関連付けアプリケーション160へ送り、テキスト関連付けアプリケーション160は、イメージ処理API145のテキストをインポートする機能呼び出すことによって変換されたテキストを開いているソースファイル内にインポートする(ステップ355)。

【0019】次に、テキスト関連付けアプリケーションは、イメージ処理API145のテキスト位置決め機能呼び出すことによってそのイメージコンポーネントに関連し変換しインポートしたテキストの位置決めを行ない、複合画像を作成する(ステップ360)。変換されたテキストの位置決め操作は、デフォルトのテキスト位置を包含することが可能であり、そのデフォルトのテキスト位置はイメージコンポーネントの下側中央部分とすることが可能である。然しながら、変換されたテキストは、任意の配向状態でイメージコンポーネントの区域内の任意の位置に配置させることが可能である。それは、又、例えば右-左-中央揃え、又は曲線に沿ってのフロー即ち流れ状態等の任意のテキスト配置特徴を使用して配置させることが可能である。変換されたテキストは、更に、イメージコンポーネントの区域外部の任意の位置に配置させることが可能であり、例えば、キャプションとしてイメージの下側に配置させるか、又は例えばマイクロソフトパワーポイントノーツ頁としてイメージから

離れた位置に配置させることも可能である。変換されたテキストは、イメージ処理アプリケーション140において使用可能なテキストフォーマットを使用してフォーマットすることが可能であり、それは、デフォルトのフォント及びデフォルトのポイント寸法を包含することが可能である。変換されたテキストは、ベクトルフォーマット又はビットマップフォーマットで、又は別のテキストコンポーネントファイルとして、イメージ処理アプリケーションテキストフォーマットで格納することが可能である。

【0020】変換されインポートしたテキストがイメージコンポーネントに関連して位置決めされた後に、テキスト関連付けアプリケーション160は、イメージ処理アプリケーションAPI145の印刷機能呼び出すことによってその変換されインポートされ位置決めされたテキストを包含するイメージコンポーネントを印刷する(ステップ370)。イメージと関連付けされている変換されたテキストの1例を図3に示してある。

【0021】上述したような図2の方法はユーザに対してステップ330及び360において入力を行なうためのプロンプトを与えるものであるが、ユーザ入力なしで自動的に動作するようにすることも可能である。自動的に動作する場合には、ステップ330及び335が取り除かれ、且つソースファイル内に含まれているか又はそれと関連しているオーディオコンポーネントはステップ340及び350において自動的に探し出され且つ変換されたテキストへ変換される。テキストがステップ355においてインポートされた後に、その変換されたテキストはデフォルトの位置決め設定を使用してステップ360において自動的に位置決めされる。変換されインポートされ位置決めされたテキストを包含するイメージコンポーネントはステップ370において印刷される。このプロセスは、ステップ300においてソースファイルの仕様においてワイルドカードを使用することにより、又はスクリプト又はバッチファイルを使用することによって複数個のソースファイルに関して自動的に実施させることが可能である。

【0022】一方、テキスト関連付けアプリケーション160は、イメージ処理アプリケーション140に対するプラグインとして実現することが可能である。この形態においては、テキスト関連付けアプリケーション160はイメージ処理アプリケーション140に対して幾つかのオプションを付加し、それはイメージをモニタ110上に表示する任意の時間においてユーザによって選択することが可能である。ユーザはソーステキストファイルを選択することによって前にオーディオから変換した変換されたテキストを表示されたイメージへ付加させることが可能であり、次いで、それに対してステップ360の方法が適用される。オーディオコンポーネントが現在表示されている画像と関連付けされている場合には、

ユーザは、ステップ350及び355の方法を使用して、オーディオコンポーネントを変換されたテキストへ変換し且つその変換されたテキストをイメージ処理アプリケーションへインポートすることを選択することが可能である。ユーザは、更に、別のオーディオファイルを選択することによって又はイメージ処理アプリケーションに対してオーディオデータを供給するための別の手段を使用することによって、オーディオを表示されたイメージと関連付けさせ、次いで、該オーディオをステップ350及び355の方法を使用して変換されたテキストへ変換させることを選択することが可能である。

【0023】イメージは、多数の手段によってそれと関連するオーディオ及び／又は変換されたテキストコンポーネントと共に保存することが可能である。変換されたテキストコンポーネントは、イメージ処理アプリケーション140のスタンダードの態様でイメージに対して直接的に取付けることが可能である。変換されたテキストコンポーネントは、イメージコンポーネントファイル内において、又は別のテキストファイル内において、オーディオコンポーネントファイルと同一のファイル内に格納することが可能である。そのファイルは、該変換されたテキストを該アプリケーションプログラムのオーディオコンポーネントに対するエイリアスとさせることによって保存することが可能であり、それは保存されたファイルの付加的なコンポーネントである。オーディオコンポーネントが変化した場合には、エイリアスを使用することは、ステップ350の方法を使用して該変換されたテキストコンポーネントの自動的なアップデートを行なうことを可能とする。イメージが複数のイメージを包含する単一ファイル内に保存されるべき場合には、該複数のイメージのうちの特定の1つと該変換されたテキストコンポーネントを関連付けさせるタグ又は識別子を形成することが可能である。

【0024】オーディオコンポーネントは多数の手段によってイメージコンポーネントと関連付けさせることが可能である。オーディオコンポーネントは、イメージ処理アプリケーション140のスタンダードの手順に従ってイメージコンポーネントと同一のファイル内に保存することが可能であり、その場合には、該オーディオコンポーネント及び該イメージコンポーネントに対するタグを使用してこれら2つのコンポーネントの間の関連を表わす。このようなタグによる方法は、例えば、フラッシュピククスフォーマットによって使用されている。

【0025】一方、オーディオコンポーネントはイメージコンポーネントが格納されているファイルとは別のファイル内に格納することが可能である。この場合には、別のオーディオファイルのファイルの名前のコンポーネントがイメージコンポーネントファイルのファイルの名前のコンポーネントとマッチングする場合がある。オーディオコンポーネントファイル内のデジタルタグは、イ

メージコンポーネントファイル内のデジタルタグとマッチングする場合がある。イメージコンポーネントファイルと関連付けされるべき1つを超えたオーディオコンポーネントが存在する場合には、同一のオーディオタグがオーディオコンポーネントファイル間において共用されている。1つ又はそれ以上のオーディオコンポーネントによって共用されている1つを超えるイメージコンポーネントが存在する場合には、同一のオーディオタグが夫々のイメージコンポーネントファイルの間で共用されている。ユーザは、手作業によってオーディオコンポーネントをイメージコンポーネントと関連付けさせることが可能である（例えば、オーディオをイメージとは独立的に採取しオーディオテープをアナログフィルムに関連付けさせる場合）。オーディオコンポーネントの一部は、イメージコンポーネントの一部に対するタグとすることが可能である（例えば、「3」という単語は、イメージコンポーネントにおける複数のイメージの3番目のイメージとマッチングするために使用することが可能である）。

【0026】オーディオデータから変換されたテキストへの変換、及びその後の変換されたテキストとイメージデータとの関連付けは、多数の手段によって実行することが可能である。オーディオデータが音声認識アプリケーション150によって認識されることのないデジタルオーディオフォーマットで格納されている場合には、そのオーディオデータは、スピーカを介して発声させるか、又はアナログオーディオ信号へ変換させ且つ適宜のプレイヤーソフトウェアを使用して出力ライン上で出力させ、次いでマイクロフォン又は入力ラインを介して適宜のデジタルオーディオフォーマットで音声認識アプリケーション150内へインポートさせることが可能である。

【0027】ドラゴンシステムのドラゴンディクテイトとすることの可能な音声認識アプリケーション150は、スタンダードのAPIを使用してデジタルオーディオを変換されたテキストへ変換させるために使用することが可能である。次いで、イメージ処理アプリケーション140は、スタンダードのテキストをインポートする技術を使用し、その変換されたテキストをインポートし且つその変換されたテキストをステップ360に従ってイメージに関連付けて位置決めさせる。その変換され位置決めされたテキストを包含するイメージは、ステップ370に従って印刷させるか、又はモニタ上で表示させるか、又は上述したようにファイル内に保存することが可能である。

【0028】イメージが例えばネガフィルム又はプリント等のデジタルイメージ以外のものである場合には、ステップ350の方法を使用してオーディオデータを変換されたテキストへ変換することが可能である。次いで、スタンダードの印刷技術を使用してその変換されたテキ

ストをイメージ上に物理的に印刷させることが可能である。

【0029】オーディオ、テキスト、イメージコンポーネントは、多数の形態をとることが可能であり、且つ以下のものに制限するわけではないが以下のものを包含する多数の手段によって作成することが可能である。

【0030】イメージは、デジタルカメラ又は例えばフィルムカメラ又はVHSレコーダ等のその他の画像形成装置によってイメージをキャプチャ即ち採取することにより作成することが可能である。イメージは、写真をデジタル化させるか、オブジェクトをスキャニングするか、又はベクトルイメージをラスター化形態へ変換させることによって作成することが可能である。

【0031】次のものに制限されるべきものではないが、使用することの可能なデジタルビットマップ及びビクセルマップイメージフォーマットとしては、グラフィックスインターチェンジフォーマット(GIF)、ジョイントフォトグラフィックエクスパートグループフォーマット(JPEG)、タグ付きイメージファイルフォーマット(TIFF)、マイクロソフトウィンドウズビットマップグラフィックスフォーマット(BMP)、アドビフォトショップフォーマット、及びフラッシュピックスフォーマット等がある。次のものに制限されるべきものではないが、使用可能なベクトルイメージとしては、ポストスクリプトファイル、アドビイラストレーターファイル、変換ビットマップイメージ等がある。次のものに制限されるべきものではないが、使用可能なアナログイメージとしては、写真フィルムイメージ(単一又はマルチフレーム、ネガ又はポジ)、及び例えばVHS等の動画ビデオイメージ等がある。使用することの可能なドキュメント頁又はドキュメント頁コンポーネントを表わすイメージとしては、アドビポータブルドキュメントフォーマット(PDF)頁又はサブ頁、文書処理ドキュメントにおけるイメージ、又はスプレッドシートの1つ又はそれ以上のセル等がある。

【0032】イメージは複数個のフレームを包含することが可能であり、その場合には、ユーザは単に最初のフレーム上の又は複数個のマルチフレーム上に関連付けされた変換されたテキストを包含させるためのオプションを提供させることが可能である。ユーザは、又、例えば、変換されたテキストの特別の部分イメージコンポーネント内のマルチフレームの各々と関連付けさせることにより、その変換され関連付けされたテキストをマルチフレームの1つ又はそれ以上を横断して分布させることを選択することが可能である。

【0033】オーディオデータは、イメージデータを作成又は採取するために使用される装置とは独立しているデジタル又はアナログオーディオレコーダを使用して作成することが可能である。例えば、イメージデータは、デジタルカメラ170によってキャプチャすることが可

能であり且つオーディオデータはマイクロフォン及びDATレコーダによってデジタルオーディオテープ(DAT)上にキャプチャすることが可能である。次いで、そのイメージデータ及びオーディオデータは、各々、スタンダードの手段を使用してコンピュータ100a内に別々にインポートされる。そのオーディオデータは、そうである必要はないが、記録した人間の音声とすることが可能である。

【0034】オーディオデータは、デジタル又はアナログ記録とすることが可能である。オーディオデータは圧縮されたもの又は圧縮されていないものとすることが可能である。デジタル記録は、例えばリアルオーディオ(Real Audio)、クイックタイム(Quick Time)、又はフラッシュピックス(Flash Pix)等のコンピュータによってサポートされているフォーマットで格納することが可能である。アナログ記録は、従来のテープ、フィルムストリップ、又はその他の媒体上に格納し、且つ例えば、スピーカを介してその記録をプレイさせることによって音声認識アプリケーション150によりオーディオテキストへ変換させ且つそのオーディオデータをコンピュータ100aのI/Oポート200a-fへ取付けられているマイクロフォンで採取するか、又はアナログオーディオ信号へ変換させ且つ適宜のプレイヤーソフトウェアを使用して出力ライン上に出力させ、次いで入力ラインを介して音声認識アプリケーション150内へインポートさせることが可能である。

【0035】イメージデータ及びそれと関連するオーディオデータは、必ずしも必要であるというわけではないが、同時に発生させるか又はキャプチャ即ち採取することが可能である。例えば、予め記録してあるサンプルからなるライブラリからのオーディオデータをイメージをキャプチャする時又は後の時のいずれかにイメージデータと関連付けさせることが可能である。ある場合には、イメージデータをキャプチャし且つ例えばコンピュータモニタ上でそのイメージデータを見る場合等の後の時間において関連するオーディオデータを記録することが有用な場合がある。

【0036】本発明は音声から派生させた変換テキストを使用するものであるが、オーディオ記録の任意のコンポーネントから変換テキストを派生させることも可能である。例えば、ステップ350において、人の歌声の記録をテキストへ変換させることが可能であり、又は複数個の音声を含むオーディオ記録から単一の音声を選択し、次いで変換テキストへ変換させることが可能である。

【0037】別の実施例においては、サブタイトル即ち説明字幕を映画の上に重ね合わせるが、その場合にオーディオデータはフィルムの元の言語のままであり、且つ変換されたテキストは別の言語である。図4を参照する

と、フィルムからのオーディオデータをキャプチャ即ち採取する(ステップ400)。そのオーディオデータは、音声認識ソフトウェア150を使用して変換されたテキストへ変換する(ステップ410)。その変換されたテキストは、例えばカリフォルニア州メンローパークのニュアンスコミュニケーションズ(Nuance Communications)及びSRIインターナショナルによって開発されたニュアンススピーチ認識エンジン(Nuance Speech Recognition Engine)のようなソフトウェアを使用して所望の説明字幕用の言語へ翻訳される。変換され翻訳されたテキストは、そのオーディオデータに関連している一連のイメージの上に重ね合わされる(ステップ430)。同様に、例えばテレビ等のビデオディスプレイ上で表示されるビデオからのオーディオデータに関連して、そのオーディオデータが変換されたテキストへ変換され且つそのビデオと共に表示され聴力障害者に対するサポートを与える。

【0038】図5を参照すると、本発明は、デジタル電子回路において、又はコンピュータハードウェア、ファームウェア、ソフトウェア又はそれらの組合わせて実現することが可能である。本発明装置は、コンピュータプロセッサによって実行するための機械読取可能な格納装置において実現したコンピュータプログラム製品の形態で実現することが可能であり、且つ本発明の方法ステップは、入力データに関して操作を行ない且つ出力を発生することにより本発明の機能を実行するためのプログラムを実行するコンピュータプロセッサによって実行することが可能である。適切なプロセッサ500は、1例としては、汎用及び特別目的の両方のマイクロプロセッサを包含している。一般的には、プロセッサは、リードオンリメモリ(ROM)510及び/又はランダムアクセスメモリ(RAM)505からCPUバス520を介して命令及びデータを受取る。コンピュータは、通常、更に、大量格納(記憶)インターフェース540を介して動作する内部ディスク545又はI/Oインターフェース530を介して動作する着脱自在のディスク535等の格納媒体からプログラム及びデータを受取る事が可能である。I/Oバス525を介してI/O装置535及び545、プロセッサ500、メモリ505、510への及びそれらからのデータの流れはI/O制御器515によって制御される。ユーザ入力、キーボード、マウス、スタイラス、マイクロフォン、トラックボール、タッチスクリーン又はその他の入力装置によって得られる。これらの要素は従来のディスクトップ又はワークステーションコンピュータ及び任意のデジタルプリントエンジン600又はマーキングエンジン、ディスプレイモニタ、又は用紙、フィルム、ディスプレイスクリーン又はその他の出力媒体上にカラー又はグレイスケールのピクセルを発生させることの可能なその他のラスタ出力装置と関

連して使用することの可能な本明細書に記載した方法を実現するコンピュータプログラムを実行するのに適したその他のコンピュータにおいて使用されるものである。

【0039】1例として、例えばポストスクリプト言語等の頁記述言語用のインタプリタを実現する印刷装置550は、プリンタランダムアクセスメモリ(RAM)580及びプリンタリードオンリメモリ(ROM)590上に格納されており且つプリンタマーキングエンジン600を制御するプログラム命令(フォント命令を含む)を実行するためのマイクロプロセッサ570を有している。RAM580は、オプションとして、例えばハードディスク(不図示)等の大量格納装置によって補充される。

【0040】コンピュータプログラム命令を実施するのに適した格納装置としては、例えばEPROM、EEPROM、フラッシュメモリ装置等の半導体メモリ装置を包含する全ての形態の非揮発性メモリ、例えば内部ハードディスク545及び着脱自在なディスク535等の磁気ディスク、MOディスク、及びCD-ROMディスク等がある。前述したものいづれの場合にも、特別に設計されたASIC(応用特定集積回路)によって補充されるものであるか又はその中に組込んだものとする事が可能である。

【0041】本発明の要素はソフトウェアとして実現された場合について説明したものであるが、本発明はソフトウェア又はハードウェア又はファームウェア又はそれら3つの組合わせて実現することが可能である。

【0042】以上、本発明の具体的実施の態様について詳細に説明したが、本発明は、これら具体例にのみ制限されるべきものではなく、本発明の技術的範囲を逸脱することなしに種々の変形が可能であることは勿論である。

【図面の簡単な説明】

【図1】 本発明に基づいて構成したシステムを示した概略ブロック図。

【図2】 テキスト関連付けアプリケーションを使用してイメージファイルを印刷するためにユーザの選択によって開始されるイベントのシーケンスを示したフローチャート。

【図3】 イメージに関連付けされている変換されたテキストを示した説明図。

【図4】 ある言語におけるオーディオデータを別の言語で説明字幕へ変換する方法を示したフローチャート。

【図5】 本発明を実現するのに適したコンピュータ及びコンピュータ要素を示した概略図。

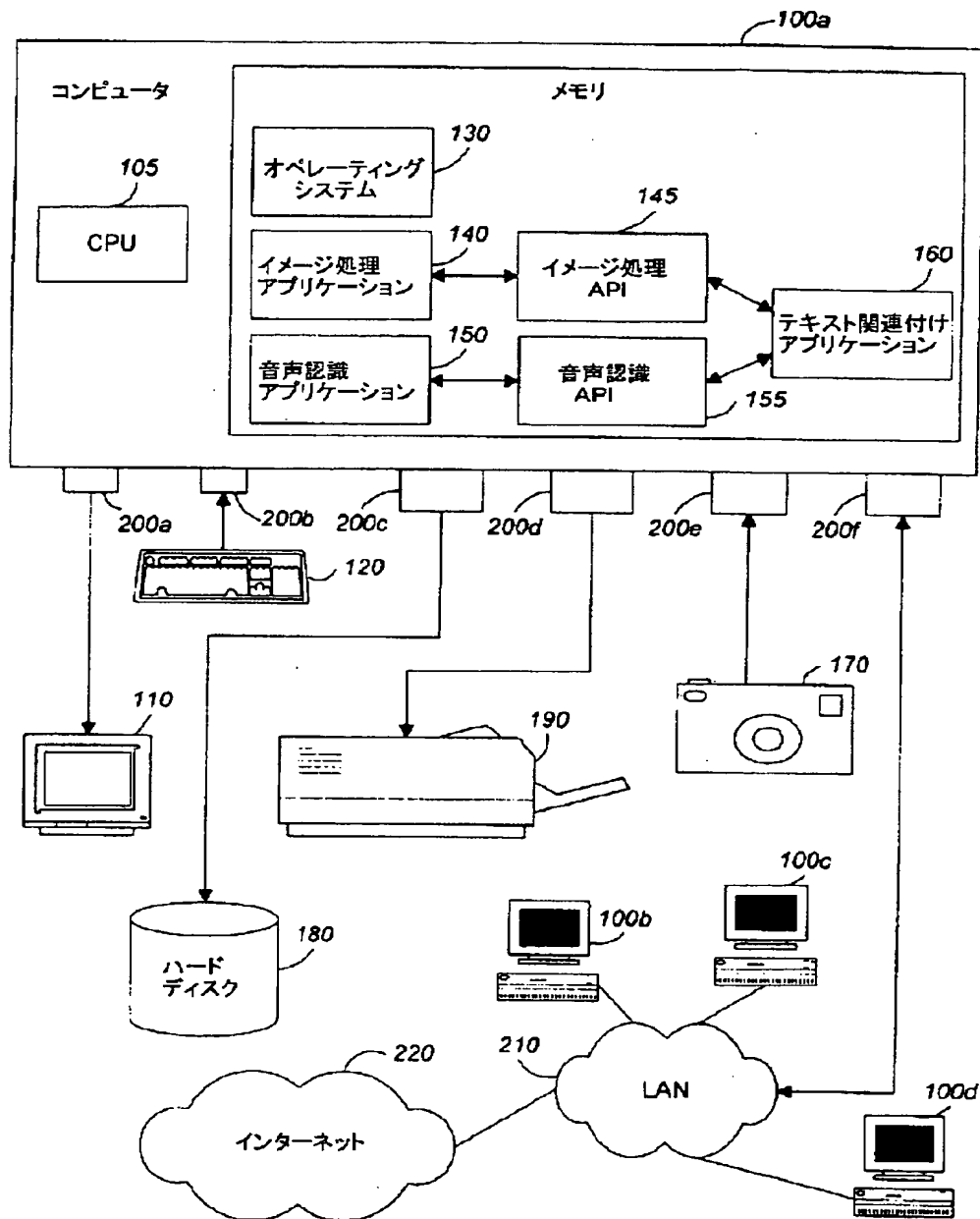
【符号の説明】

- 100a コンピュータ
- 105 CPU
- 110 モニタ
- 120 キーボード

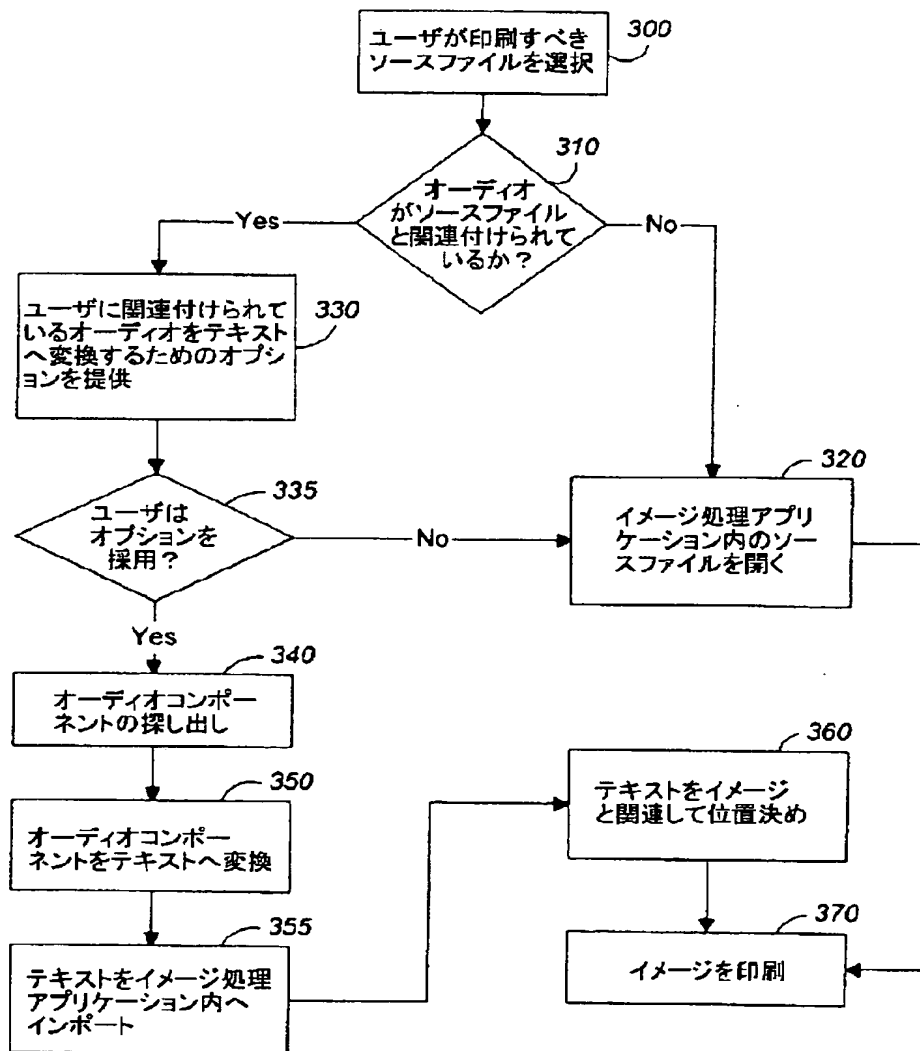
130 オペレーティングシステム
 140 イメージ処理アプリケーション
 145 イメージ処理API
 150 音声認識アプリケーション
 155 音声認識API

160 テキスト関連付けアプリケーション
 170 デジタルカメラ
 180 ハードディスク
 190 プリンタ
 200 I/Oポート

【図1】



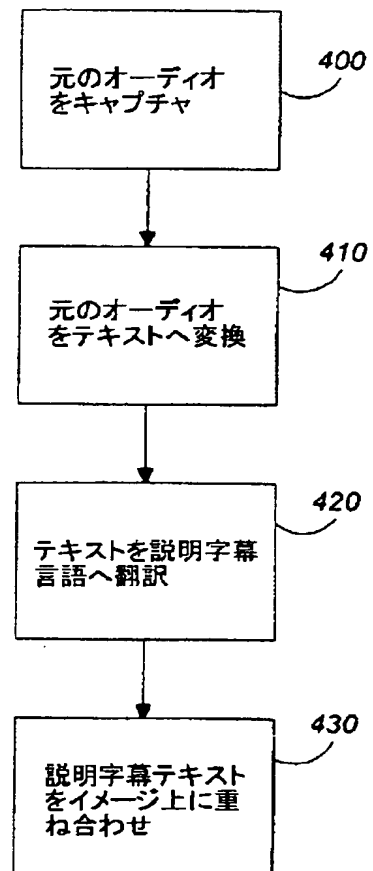
【図2】



【図3】



【図4】



【図5】

